



TITLE:

# 再生核ヒルベルト空間を用いた統計的推論 (再生核の応用についての研究)

AUTHOR(S):

福水, 健次

---

CITATION:

福水, 健次. 再生核ヒルベルト空間を用いた統計的推論 (再生核の応用についての研究). 数理解析研究所講究録 2008, 1618: 41-63

ISSUE DATE:

2008-12

URL:

<http://hdl.handle.net/2433/140195>

RIGHT:

## 再生核ヒルベルト空間を用いた統計的推論

By

福水 健次\*

### §1. はじめに

機械学習ないしは統計的学習の分野では、1990 年代半ばごろからサポートベクターマシン (SVM) (Boser et al., 1992; Vapnik, 1998) と呼ばれるパターン識別の方法が注目を集め、文字認識などさまざまな応用において高い能力を発揮することが示された<sup>1</sup>。SVM の特徴のひとつとして、正定値カーネルによってデータを変換することによりデータの非線形特徴を取り込んだ点が挙げられる。SVM の成功を受けて、正定値カーネルないしは再生核ヒルベルト空間を用いたデータ解析の方法論である「カーネル法」(Schölkopf and Smola, 2002) の研究が盛んとなった。

カーネル法の利点は、再生核ヒルベルト空間を用いることにより、線形の関係に基づく古典的なデータ解析手法を、計算をあまり複雑にすることなく非線形に拡張可能な点にある。一方最近になって、データを再生核ヒルベルト空間内の関数データに変換することによって、独立性、条件付独立性、分布の同一性といった確率変数の統計的性質を議論するための新しい組織的な方法が提案されてきた (Bach and Jordan, 2002; Fukumizu et al., 2004; Gretton et al., 2007; Fukumizu et al., 2008b)。

本稿は、カーネル法の一般的な解説をごく簡単に行った後、再生核ヒルベルト空間を用いた確率分布に関する推論についての筆者らの最近の研究について解説する。

以下では、 $\mathbb{R}$  と  $\mathbb{C}$  によりそれぞれ実数全体および複素数全体をあらわし、 $\mathbb{R}^m$  は  $m$  次元ユークリッド空間をあらわす。

---

Communicated by JJ, June 8, 2000 Revised September 11, 2000.

2000 Mathematics Subject Classification(s): 46E22, 62H20, 62H99

本研究の一部は科研費基盤研究 (C) 19500249 および Alexander von Humboldt 財団フェローシップにより支援を受けた。

\*統計数理研究所 〒106-8569 東京都港区南麻布 4-6-7.

<sup>1</sup>本稿では SVM は議論しないので、興味ある読者は Schölkopf and Smola (2002) などを見ていただきたい。

## §2. 再生核ヒルベルト空間によるデータ解析

一言で述べるならば、カーネル法はデータの変換を行う方法論である。データに何らかの変換を施してから解析する手法は古くから存在するが、カーネル法の利点は、再生核ヒルベルト空間への変換を用いることにより、変換後のデータに対する線形のデータ解析が比較的少ない計算量で行える点にある。本稿ではカーネル法のさまざまな手法に関しては詳しく述べないので、より詳しく知りたい読者は Schölkopf and Smola (2002) などの教科書を見ていただきたい。

### §2.1. 正定値カーネルと再生核ヒルベルト空間

以下では特に断らない限り、正定値カーネルは実数値とし、再生核ヒルベルト空間は実ヒルベルト空間とする。また、再生核ヒルベルト空間はすべて可分とする。

集合  $\Omega$  上の正定値カーネル  $k$  の定める再生核ヒルベルト空間を  $\mathcal{H}$  とするとき、よく知られているように (i) 任意の  $x \in \Omega$  に対して  $k(\cdot, x) \in \mathcal{H}$  であり、 $\{k(\cdot, x) \in \mathcal{H} \mid x \in \Omega\}$  の張る線形空間は  $\mathcal{H}$  で稠密 (ii) 任意の  $f \in \mathcal{H}$  と  $x \in \Omega$  に対し、再生性

$$(1) \quad \langle f, k(\cdot, x) \rangle_{\mathcal{H}} = f(x)$$

が成り立つ。ここで  $\langle \cdot, \cdot \rangle_{\mathcal{H}}$  は  $\mathcal{H}$  の内積を表す。

データ解析においては、(ii) の再生性により、ヒルベルト空間  $\mathcal{H}$  での内積計算が数値的に容易に実行可能となる点が重要である。例えば、 $x_1, \dots, x_n \in \Omega$  に対し、 $f = \sum_{i=1}^n a_i k(\cdot, x_i)$ ,  $g = \sum_{j=1}^n b_j k(\cdot, x_j)$  という2つの  $\mathcal{H}$  の元の内積は

$$\langle f, g \rangle_{\mathcal{H}} = \sum_{i=1}^n \sum_{j=1}^n a_i b_j k(x_i, x_j)$$

で与えられ、 $k$  の値の評価に還元される。 $n$  次半正定値行列  $(k(x_i, x_j))_{i,j=1}^n$  はグラム行列と呼ばれる。

次に、正定値カーネルおよび再生核ヒルベルト空間をデータ解析に用いる方法について述べる。統計的なデータ解析においては、データは確率的に発生するものと考え確率変数により表現する。 $(\Omega, \mathcal{B})$  を可測空間とし、 $X_1, \dots, X_n$  を  $\Omega$  上の  $n$  個のデータ、すなわち  $\Omega$  に値をもつ確率変数とする。 $k$  を  $\Omega$  上の可測な正定値カーネル、 $\mathcal{H}$  を  $k$  が定める再生核ヒルベルト空間とし、写像  $\Phi$  を

$$(2) \quad \Phi: \Omega \rightarrow \mathcal{H}, \quad x \mapsto k(\cdot, x)$$

により定義する。データ  $X_i$  を  $\Phi$  により変換することによって、 $\mathcal{H}$  上のデータ  $\{\Phi(X_i)\}_{i=1}^n = \{k(\cdot, X_i)\}_{i=1}^n$  が得られる。 $\mathcal{H}$  は関数空間なので  $\Phi(X_i)$  を関数データと呼ぶこともある。

カーネル法の要点は、もとのデータ  $X_1, \dots, X_n$  のかわりに  $\mathcal{H}$  内の関数データ  $\Phi(X_1), \dots, \Phi(X_n)$  を用いるところにある。 $\mathcal{H}$  は内積を持つベクトル空間であるので、 $\mathbb{R}^m$  値の確率変数に対する通常のデータ解析手法の多くを拡張することが可能である。写像  $\Phi$  は一般に非線形写像であるので、 $\Phi(X)$  はデータの非線形特徴を表し、カーネル法は非線形データ解析と考えることができる。このとき、再生核ヒルベルト空間を用いると、下の例からもわかるように、データ解析に必要な計算が実用的な範囲に収まる場合が多い点が有利である。この方法論は線形手法の「カーネル化」と呼ばれ、主成分分析、Fisher 判別分析、正準相関分析など様々な古典的データ解析手法のカーネル化が可能である。また、 $\Omega$  は必ずしもユークリッド空間である必要はなく、正定値カーネルが定義できれば任意の測度集合に対して同じ方法論が適用できる。これにより、カーネル法はグラフやシンボル列など非ユークリッド的データに対するデータ解析にも適用しやすい利点を持つ。

**カーネル法の例：カーネル主成分分析** 典型的な例として、カーネル主成分分析 (Kernel PCA, Schölkopf et al., 1998) を簡単に紹介する。主成分分析は、 $\mathbb{R}^m$  上に与えられたデータを低次元部分空間で表現する方法であり、データを用いた経験分散共分散行列の大きい固有値に対応する固有ベクトルの張る部分空間に射影することにより表現を得る。カーネル主成分分析は、その再生核ヒルベルト空間への拡張である。

式 (2) の変換によって得られたデータ  $\Phi(X_1), \dots, \Phi(X_n)$  に対し、再生核ヒルベルト空間  $\mathcal{H}$  上で主成分分析を行うと、その第 1 主成分の軸は以下の最大値問題の解によって与えられる：

$$(3) \quad \max_{f \in \mathcal{H}, \|f\|_{\mathcal{H}}=1} \sum_{i=1}^n \langle f, \Phi(X_i) - \hat{m} \rangle_{\mathcal{H}}^2$$

ここで、 $\hat{m} \in \mathcal{H}$  は  $\{\Phi(X_i)\}_{i=1}^n$  のサンプル平均

$$(4) \quad \hat{m} = \frac{1}{n} \sum_{i=1}^n \Phi(x_i) = \frac{1}{n} \sum_{i=1}^n k(\cdot, X_i)$$

であり、 $\|f\|_{\mathcal{H}}$  は  $\mathcal{H}$  のノルムを表す。

式 (3) は一般に無限次元空間上の最適化となるが、この右辺は  $\{\Phi(x_1) - \hat{m}, \dots, \Phi(x_n) - \hat{m}\}$  の張る高々  $n-1$  次元部分空間の直交補空間に依存しないため、

$$f = \sum_{j=1}^n a_j (\Phi(X_j) - \hat{m})$$

( $a_j$  は実数) という形で軸  $f$  を探せば十分である。このとき式 (3) は

$$\max_{a \in \mathbb{R}^n, a^T G a = 1} a^T G^2 a$$

となり、固有値問題として容易に最適解が得られる。ここで、 $n \times n$  行列  $G$  は中心化グラム行列と呼ばれる対称行列で、

$$(5) \quad G_{ij} = k(x_i, x_j) - \frac{1}{n} \sum_{b=1}^n k(x_i, x_b) - \frac{1}{n} \sum_{a=1}^n k(x_a, x_j) + \frac{1}{n^2} \sum_{a=1}^n \sum_{b=1}^n k(x_a, x_b)$$

により定義される。

以上の導出は、カーネル法の方法論の典型的なものである。この方法論では、変換された  $\mathcal{H}$  内の関数データに対し、データの内積を基本とした  $\mathbb{R}^m$  上での解析手法を適用する。解  $f$  は一般には無限次元空間の元であるが、与えられた  $n$  個の関数データ  $\Phi(x_i)$  の張る有限次元部分空間内で解を求めれば十分であることが多く (Representer 定理, Schölkopf and Smola (2002), Section 4.2)、その範囲で問題を書き直すと  $n$  次のグラム行列を用いた問題に還元される。

ユークリッド空間  $\mathbb{R}^m$  上の正定値カーネルのなかでデータ解析によく用いられる例は、多項式カーネル

$$k_{d,c}^{poly}(x, y) = (x^T y + c)^d \quad (c \geq 0, d \in \mathbb{N}),$$

ガウス RBF(Radial Basis Function) カーネル

$$k_{\sigma}^G(x, y) = \exp\left(-\frac{\|x - y\|^2}{2\sigma^2}\right), \quad (\sigma > 0),$$

ラプラスカーネル

$$k_{\beta}^L(x, y) = \exp(-\beta \sum_{i=1}^m |x_i - y_i|) \quad (\beta > 0)$$

などである。多項式カーネル  $k_{d,c}^{poly}$  ( $c > 0$ ) の定める再生核ヒルベルト空間は、ベクトル空間として  $d$  次以下の多項式全体と一致する。ガウス RBF カーネルとラプラスカーネルが定める再生核ヒルベルト空間は無限次元となる。

データに変換を施すことにより非線性を取り込む手法は古くからある。例えば、実数値データ  $X$  を  $(X, X^2, X^3, \dots)$  と冪により拡張した後に解析を行うことにより高次モーメントを取り込むことは可能である。しかしながら、例えば  $m$  次元データを  $d$  次までの冪によって拡張すると、 $m$  の増加に対して次元が増大し、その後の解析の計算量に問題が生じる。例えば、200 次元のデータに対して 3 次までの冪変換を施すと、131 万以上の次元のデータを扱う必要がある。正定値カーネルを用いると、直接的な展開による内積計算を避けることができ、データ数の次元の計算で済む。したがって、高次元でデータ数が比較的少ない場合には計算上有利である。逆にデータ数が大きいと、計算量を実用的なレベルにまで引き下げる工夫が必要となる。

### §3. 再生核ヒルベルト空間による確率分布に関する推論

以下では、再生核ヒルベルト空間への変換を用いて、確率変数の同一性、独立性、条件付独立性などの性質を推論する方法に関して述べる。以降では、測度空間  $(\Omega, \mathcal{B}, \mu)$  に対して 2 乗可積分な実関数全体のなすヒルベルト空間を  $L^2(\Omega, \mu)$  で表し、紛れがない場合は  $L^2(\mu)$  と略記する。

#### §3.1. 再生核ヒルベルト空間における平均と共分散

ヒルベルト空間  $\mathcal{H}$  はボレル集合族によって可測空間と考えることにする。 $\mathcal{H}$  に値をとる確率変数  $F$  が強可積分、すなわち  $E[\|F\|] < \infty$  であるとき、 $F$  の平均  $E[F] \in \mathcal{H}$  が定まる。これを  $m_F$  で表す。このとき

$$(6) \quad E[\langle F, f \rangle_{\mathcal{H}}] = \langle m_F, f \rangle_{\mathcal{H}} \quad (\forall f \in \mathcal{H})$$

が成り立つ。Riesz の補題により、関係式 (6) を平均  $m_F$  の定義としてもよい。

いま、 $(\mathcal{X}, \mathcal{B})$  を可測空間、 $X$  を  $\mathcal{X}$  に値をとる確率変数、 $\mathcal{H}$  を  $\mathcal{X}$  上の可測な正定値カーネル  $k$  の定める再生核ヒルベルト空間とする。以下では常に、

$$(A-1) \quad E[k(X, X)] < \infty$$

という仮定をおく。

前章と同様、写像  $\Phi: \mathcal{X} \rightarrow \mathcal{H}$  を  $\Phi(x) = k(x, \cdot)$  により定めると、 $\|\Phi(X)\|_{\mathcal{H}}^2 = k(X, X)$  により、仮定 (A-1) は、 $\Phi(X)$  が 2 次強可積分な  $\mathcal{H}$  値確率変数であることと同値である。

$\Phi(X)$  の平均を  $m_X$  とおくと、式 (6) より

$$(7) \quad \langle f, m_X \rangle_{\mathcal{H}} = E[f(X)] \quad (\forall f \in \mathcal{H})$$

が成立する。 $m_X \in \mathcal{H}$  を  $X$  の  $\mathcal{H}$  における平均と呼ぶ。 $X$  の分布が  $P$  であるとき、その平均を  $m_P$  と書くこともある。式 (7) において  $f = k(u, \cdot)$  とおくと、

$$m_X(u) = E[k(u, X)]$$

となり、関数  $m_X$  の陽な表現を得る。

$X$  が  $\mathbb{R}^m$  値の確率ベクトルの場合、非線形な正定値カーネル  $k$  を用いると、平均  $m_X$  は  $X$  の分布の高次モーメントの情報を持つ。例えば、2 節で述べたように、 $d$  次の多項式カーネル ( $c > 0$ ) が定める再生核ヒルベルト空間  $\mathcal{H}$  はベクトル空間として  $d$  次以下の多項式全体と一致するので、 $\mathbb{R}$  上の確率変数  $X$  に対し、その  $r$  次モーメント  $m_r = E[X^r]$  ( $0 \leq r \leq d$ ) が

$$m_r = \langle X^r, m_X \rangle_{\mathcal{H}}$$

により計算される。

次に再生核ヒルベルト空間上の相互共分散作用素を定義する。 $(\mathcal{X}, B_X), (\mathcal{Y}, B_Y)$  を可測空間、 $(X, Y)$  は  $\mathcal{X} \times \mathcal{Y}$  に値をとる確率変数とする。 $k_X, k_Y$  をそれぞれ  $\mathcal{X}, \mathcal{Y}$  上の可測な正定値カーネルとし、 $\mathcal{H}_X, \mathcal{H}_Y$  をそれぞれ  $k_X, k_Y$  の定める再生核ヒルベルト空間とする。ここで、確率変数と正定値カーネルは (A-1) の仮定を満たすとする。

$\mathcal{H}_X \otimes \mathcal{H}_Y$  により  $\mathcal{H}_X$  と  $\mathcal{H}_Y$  の直積<sup>2</sup>をあらわす。平均の場合と同様に、 $\Phi_X(x) = k_X(\cdot, x), \Phi_Y(y) = k_Y(\cdot, y)$  を定義し、 $\mathcal{H}_X \otimes \mathcal{H}_Y$  上の確率変数  $G_{XY} = (\Phi_X(X) - m_X) \otimes (\Phi_Y(Y) - m_Y)$  の平均を  $m_G$  とする。このとき平均の性質 (7) から、任意の  $f \in \mathcal{H}_X, g \in \mathcal{H}_Y$  に対して

$$\langle m_G, f \otimes g \rangle_{\mathcal{H}_X \otimes \mathcal{H}_Y} = E[f(X)g(Y)] - E[f(X)]E[g(Y)] = \text{Cov}[f(X), g(Y)]$$

が成り立つ。 $m_G \in \mathcal{H}_1 \otimes \mathcal{H}_2$  を  $\mathcal{H}_1$  から  $\mathcal{H}_2$  への作用素と同一視し、これを  $\Sigma_{YX} : \mathcal{H}_X \rightarrow \mathcal{H}_Y$  とかくと、

$$(8) \quad \langle g, \Sigma_{YX} f \rangle_{\mathcal{H}_Y} = \text{Cov}[f(X), g(Y)]$$

が成立し、 $f(X)$  と  $g(Y)$  の共分散をあらわす。 $\Sigma_{YX}$  は有界線形作用素であり、 $(X$  と  $Y$  の定める) **相互共分散作用素** と呼ばれる<sup>3</sup>。特に  $Y = X$  の場合  $\Sigma_{XX}$  は自己共役作用素で、**共分散作用素** と呼ばれる。

相互共分散作用素は以下のような分解を持つ (Baker, 1973)。一般に、相互共分散作用素  $\Sigma_{YX}$  に対し、作用素ノルムが 1 以下の有界作用素  $V_{YX} : \mathcal{H}_X \rightarrow \mathcal{H}_Y$  が一意的に存在し

$$\Sigma_{YX} = \Sigma_{YY}^{1/2} V_{YX} \Sigma_{XX}^{1/2}$$

かつ  $\mathcal{R}(V_{YX}) \subset \overline{\mathcal{R}(\Sigma_{YY})}, \mathcal{N}(V_{YX})^\perp \subset \overline{\mathcal{R}(\Sigma_{XX})}$  となる。ここで  $\mathcal{N}(A)$  と  $\mathcal{R}(A)$  はそれぞれ作用素  $A$  の零核と値域を表す。

### §3.2. カーネル法による分布の特徴づけ

前節で見たように、確率変数を再生核ヒルベルト空間に写像すると、その平均はさまざまなモーメントの情報を含んでいる。大雑把に言えば、確率変数に対してすべてのモーメントを考えればその分布は決まるので、十分広いクラスの関数を含むような再生核ヒルベルト空間における平均を考えれば、確率変数を一意的に定めることが期待できる。そこでまず特性的な正定値カーネルを定義する。

#### 特性的な正定値カーネル

<sup>2</sup> 積  $k_X k_Y$  により定まる  $\mathcal{X} \times \mathcal{Y}$  上の再生核ヒルベルト空間を  $\mathcal{H}_X$  と  $\mathcal{H}_Y$  の直積といい、 $\mathcal{H}_X \otimes \mathcal{H}_Y$  で表す。Aronszajn (1950) 参照。

<sup>3</sup> 相互共分散作用素の一般的な理論は Baker (1973) に詳しい。

$(\mathcal{X}, \mathcal{B}_\mathcal{X})$  を可測空間、 $\mathcal{P}$  をその上の確率測度全体の族とする。 $\mathcal{X}$  上の有界かつ可測な正定値カーネル  $k$  が**特性的** (characteristic) であるとは、写像

$$\mathcal{P} \rightarrow \mathcal{H}_k, \quad P \mapsto m_P$$

が単写であることをいう。すなわち、 $k$  の定める再生核ヒルベルト空間を  $\mathcal{H}$  とするとき、 $\int f dP = \int f dQ$  ( $\forall f \in \mathcal{H}$ ) ならば  $P = Q$  であることをいう。言い換えると、特性的なカーネルは、関数  $E[k(X, u)]$  により  $X$  の確率分布を一意に定めるクラスである。

$\mathbb{R}^m$  値の確率ベクトル  $X$  の特性関数は

$$E[e^{\sqrt{-1}u^T X}]$$

により定義された。よく知られているように特性関数は  $X$  の分布を一意に定める。特性的であるという性質は、確率分布を特徴付けるという特性関数の性質の一種の拡張と考えられる。

次の事実は特性的な正定値カーネルに関して基本的である。

**補題 3.1** Fukumizu et al. (2008a). 上の記法のもと、有界かつ可測な正定値カーネル  $k$  が特性的であることの必要十分条件は、任意の確率分布  $P \in \mathcal{P}$  に対して  $\mathcal{H} + \mathbb{R}$  が  $L^2(P)$  で稠密であることである。ここで、 $\mathcal{H} + \mathbb{R}$  は再生核ヒルベルト空間としての直和、すなわち  $f + c$  ( $f \in \mathcal{H}, c \in \mathbb{R}$ ) の形の関数からなる再生核ヒルベルト空間を意味する。

**証明.** まず十分性を示す。 $P, Q \in \mathcal{P}$  に対し  $m_P = m_Q$  とする。 $P \neq Q$  と仮定しよう。すると  $P - Q$  の全変動  $|P - Q|$  に対して、 $\mathcal{H} + \mathbb{R}$  は  $L^2(|P - Q|)$  で稠密なので、任意の  $\varepsilon > 0$  と  $A \in \mathcal{B}_\mathcal{X}$  に対して  $f \in \mathcal{H} + \mathbb{R}$  があって  $\int |f - I_A| d(|P - Q|) < \varepsilon$  とできる。このとき

$$|(E_P[f(X)] - P(A)) - (E_Q[f(X)] - Q(A))| < \varepsilon$$

であるが、 $m_P = m_Q$  より  $|P(A) - Q(A)| < \varepsilon$  を得る。 $\varepsilon$  は任意なので、 $P \neq Q$  に矛盾する。したがって  $P = Q$  であり、 $k$  は特性的である。

次に  $k$  を特性的とする。いま、ある  $P \in \mathcal{P}$  があって  $\mathcal{H} + \mathbb{R}$  が  $L^2(P)$  で稠密でないと仮定しよう。このとき、ある  $f \in L^2(P)$  ( $f \neq 0$ ) があって、任意の  $\varphi \in \mathcal{H}$  に対し  $\int f \varphi dP = 0$  かつ  $\int f dP = 0$  が成り立つ。 $c = \frac{1}{\|f\|_{L^1(P)}}$  として、確率測度  $Q_1, Q_2$  を

$$Q_1(E) = c \int_E |f| dP, \quad Q_2(E) = c \int_E (|f| - f) dP$$

により定義すると、 $f \neq 0$  により  $Q_1 \neq Q_2$  であるが、

$$E_{Q_1}[k(\cdot, X)] - E_{Q_2}[k(\cdot, X)] = c \int f(x) k(\cdot, x) dP(x) = 0$$



となる。これは  $k$  が特性的であることに矛盾する。  $\square$

$\mathbb{R}^m$  上の連続で平行移動不変な正定値カーネルに対しては、特性的なカーネルの特徴づけが可能である。ここで、 $\mathbb{R}^m$  上の平行移動不変な正定値カーネルとは、正定値関数  $\phi$  があって

$$(9) \quad k(x, y) = \phi(x - y)$$

と表される正定値カーネルのことである。次の結果はよく知られている。

**定理 3.1** Bochner の定理.  $\mathbb{R}^m$  上の複素数値連続関数  $\phi(z)$  に対し、 $k(x, y) = \phi(x - y)$  が複素数値正定値カーネルとなる必要十分条件は、 $\mathbb{R}^m$  上の有限な非負ボレル測度  $\Lambda$  があって、

$$(10) \quad \phi(z) = \int e^{\sqrt{-1}u^T z} d\Lambda(u)$$

と表されることである。

**定理 3.2** Sriperumbudur et al. (2008).  $\mathbb{R}^m$  上の連続かつ平行移動不変な実数値正定値カーネル  $k(x, y) = \phi(x - y)$  が  $\mathbb{R}^m$  上の有限な非負ボレル測度  $\Lambda$  によって式 (10) で表されているとする。このとき、 $k$  が特性的であるための必要十分条件は  $\text{Supp}(\Lambda) = \mathbb{R}^m$  が成り立つことである。ここで有限測度  $\mu$  に対して  $\text{Supp}(\mu) = \{x \in \mathbb{R}^m \mid x \in U \text{ なる任意の開集合 } U \text{ に対して } \mu(U) > 0\}$  と定義する。

証明. まず、平行移動不変な正定値カーネルと  $\mathbb{R}^m$  上のボレル確率測度  $P$  に対し、 $m_P(u) = E_{X \sim P}[\phi(u - X)] = P * \phi$  に注意する。ここで  $\mu * \phi$  は有限なボレル測度  $\mu$  と  $\phi$  の畳み込みをあらわす。このことから、十分性を示すためには  $\mu * \phi = 0$  ならば  $\mu = 0$  をいえばよい。以下、有限な測度  $\mu$  の Fourier 変換を  $\hat{\mu}$  で表す。Fubini の定理から

$$\begin{aligned} \int (\mu * \phi)(x) d\mu(x) &= \int \int \phi(x - y) d\mu(y) d\mu(x) \\ &= \int \int \int e^{\sqrt{-1}(x-y)^T u} d\Lambda(u) d\mu(y) d\mu(x) \\ &= \int \int e^{\sqrt{-1}x^T u} d\mu(x) \int e^{-\sqrt{-1}y^T u} d\mu(y) d\Lambda(u) = \int |\hat{\mu}(u)|^2 d\Lambda(u) \end{aligned}$$

を得る。 $\mu * \phi = 0$  のとき上式は 0 であるが、 $\hat{\mu}$  の連続性と  $\text{supp}(\Lambda) = \mathbb{R}^m$  により、 $\hat{\mu} = 0$ 、すなわち  $\mu = 0$  を得る。

$k$  が特性的であるとき、 $\text{Supp}(\Lambda) \neq \mathbb{R}^m$  としよう。 $\phi$  が実数値であることから  $\text{Supp}(\Lambda)$  が原点对称であることに注意すると、零でない  $C^\infty$  関数  $g$  で、

$\text{Supp}(g) \cap \text{Supp}(\Lambda) = \emptyset$ ,  $g(-x) = g(x)$ , かつ  $g(0) = 0$  なるものが存在する。このとき、 $g$  の逆 Fourier 変換  $\check{g}$  は実数値かつ急減少関数 (Schwartz クラス) である。したがって、 $c = 1/\int |\check{g}(x)|dx$  として

$$P(E) = c \int_E |\check{g}(x)|dx, \quad Q(E) = c \int_E (|\check{g}(x)| - \check{g}(x))dx$$

とおくと、 $\int \check{g}(x)dx = g(0) = 0$  により、 $P, Q$  はともにボレル確率測度であり、 $P \neq Q$  を満たす。一方

$$((P - Q) * \phi)^\wedge = c(\check{g} * \phi)^\wedge = cg\Lambda = 0$$

により  $m_P = m_Q$  であるので、これは  $k$  が特性的であることに矛盾する。  $\square$

上の定理により、ガウス RBF カーネル  $k_\sigma^G(x, y) = \exp\{-\|x - y\|^2/(2\sigma^2)\}$  やラプラスカーネル  $k_\lambda^L(x, y) = \exp(-\lambda \sum_{i=1}^m |x_i - y_i|)$  が  $\mathbb{R}^m$  上の特性的なカーネルであることがわかる。実際、 $\exp\{-\|x - y\|^2/(2\sigma^2)\}$  と  $\exp\{-\lambda \sum_{i=1}^m |x_i - y_i|\}$  のフーリエ変換は、正の定数倍を除いて、それぞれ  $\exp\{-\sigma^2\|x - y\|^2/2\}$  と  $\prod_{i=1}^m 1/(u_i^2 + \lambda^2)$  である。

### 分布の均一性検定への応用

特性的な正定値カーネルを用いると、平均  $m_X$  によって確率分布が識別可能である。このことを利用して、Gretton et al. (2007) では分布の均一性検定に  $m_X$  の推定量を用いる方法を提案している。以下に概要を簡単に紹介する。

分布の均一性検定とは、 $(X_1, \dots, X_n)$  と  $(Y_1, \dots, Y_n)$  という2つのサンプルを用いて、それらの分布が同一かどうかを判定する問題である。一般に2つのサンプルの大きさは異なってもよいが、ここでは簡単のため同じとする。以下では  $X_1, \dots, X_n$  と  $Y_1, \dots, Y_n$  はそれぞれ可測空間  $(\mathcal{X}, \mathcal{B})$  上の確率分布  $P$  および  $Q$  に独立に従う i.i.d. サンプルとし、 $P = Q$  を帰無仮説、 $P \neq Q$  を対立仮説として検定を行うことを考える。

$k$  を  $\mathcal{X}$  上の特性的な正定値カーネル、 $k$  が定める再生核ヒルベルト空間を  $\mathcal{H}$ 、 $P$  および  $Q$  による  $\mathcal{H}$  での平均を  $m_P, m_Q$  とし、

$$M(P, Q) \equiv \|m_P - m_Q\|_{\mathcal{H}}^2$$

と定義すると、 $P = Q$  と  $M(P, Q) = 0$  は同値である。式 (7) より

$$(11) \quad M(P, Q) = E[k(X, \tilde{X})] - 2E[k(X, Y)] + E[k(Y, \tilde{Y})]$$

と計算できる。ここで  $X, \tilde{X}$  は  $P$  に、 $Y, \tilde{Y}$  は  $Q$  に従う確率変数であり、すべて互いに独立とする。 $m_P$  および  $m_Q$  の推定量は、式 (4) と同様に

$$(12) \quad \hat{m}_P = \frac{1}{n} \sum_{i=1}^n k(\cdot, X_i), \quad \hat{m}_Q = \frac{1}{n} \sum_{i=1}^n k(\cdot, Y_i)$$

で与えられることから、検定統計量として

$$\hat{M}_n(P, Q) = \|\hat{m}_P - \hat{m}_Q\|_{\mathcal{H}}^2$$

を用いることができる。これを展開すると

$$(13) \quad \hat{M}_n(P, Q) = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=i}^n \{k(X_i, X_j) - 2k(X_i, Y_j) + k(Y_i, Y_j)\}$$

である。

帰無仮説  $P = Q$  が成り立てば、 $\hat{M}_n(P, Q)$  は  $n \rightarrow \infty$  のとき 0 に確率収束し、有限の  $n$  においても小さい値をとることが期待されるが、検定を行う際にはその分布（帰無分布）が必要になる。 $n \rightarrow \infty$  とした漸近的な状況で  $n\hat{M}_n(P, Q)$  が収束する分布は知られており (Gretton et al., 2007)、それに基づいて棄却域を決めることが可能である。また、並べ替え検定と呼ばれる手法で帰無分布をシミュレーションする方法を用いることもできる。この検定に関する詳細や具体例は Gretton et al. (2007) および福水 (2008) を見ていただきたい。

### §3.3. カーネル法による独立性の特徴づけ

2つの  $\mathbb{R}^m$  値確率ベクトル  $X, Y$  の関係を見るために共分散や相関を調べることは基本的な方法であるが、それでは線形の関係しか考慮できない。再生核ヒルベルト空間への写像  $\Phi_X(X), \Phi_Y(Y)$  は高次の情報を含んでいるため、関数空間上で相関や共分散を考えれば、確率変数の独立性や依存性が調べられることが期待できる。以下ではこの考えに基づいて、3.1 節で定義した再生核ヒルベルト空間上の相互共分散作用素を用いて確率変数の独立性や依存性をはかる方法を議論する。

まず、相互共分散作用素は以下のように確率変数の独立性を特徴付ける。

**定理 3.3.**  $(\mathcal{H}_1, k_1), (\mathcal{H}_2, k_2)$  をそれぞれ  $\mathcal{X}, \mathcal{Y}$  上の再生核ヒルベルト空間とし、積  $k_X k_Y$  は  $\mathcal{X} \times \mathcal{Y}$  上特性的であるとする。 $(X, Y)$  を  $\mathcal{X} \times \mathcal{Y}$  上の確率変数とし、それぞれ (A-1) の条件を仮定するとき、

$$(14) \quad X \perp\!\!\!\perp Y \iff \Sigma_{XY} = O \quad (\text{零作用素})$$

の同値関係が成り立つ。ここで  $X \perp\!\!\!\perp Y$  は  $X$  と  $Y$  が独立であることを表す。

**証明.**  $(X, Y)$  の同時分布を  $P_{XY}$ 、また  $X, Y$  と同じ周辺分布を持ち、互いに独立な確率分布を  $P_X \otimes P_Y$  と書く。 $\Sigma_{YX}$  は、 $\mathcal{H}_X \otimes \mathcal{H}_Y$  上の平均として

$$(15) \quad \Sigma_{YX} = m_{P_{XY}} - m_{P_X \otimes P_Y}$$

とみなせるので、定理の同値性は  $k_X k_Y$  が特性的なことから従う。  $\square$

定理 3.3 は、 $\mathbb{R}^m$  値確率ベクトルに対する特性関数を用いた独立性の特徴づけ

$$X \perp\!\!\!\perp Y \iff E_{XY}[e^{\sqrt{-1}u^T X} e^{\sqrt{-1}v^T Y}] = E_X[e^{\sqrt{-1}u^T X}] E_Y[e^{\sqrt{-1}v^T Y}]$$

のある種の一般化とみなすことができる。式 (14) の右式は  $k(u, X)$  と  $k(v, Y)$  の共分散が 0 であることと同値である。

平均の場合と同様、サンプル  $(X_1, Y_1), \dots, (X_n, Y_n)$  が与えられたとき、確率分布を経験分布に置き換えることにより経験相互共分散作用素

$$(16) \quad \hat{\Sigma}_{YX}^{(n)} = \frac{1}{n} \sum_{i=1}^n (k_Y(\cdot, Y_i) - \hat{m}_Y) \langle (k_X(\cdot, X_i) - \hat{m}_X), \cdot \rangle_{\mathcal{H}_X}$$

が得られる。

式 (15) の表現から、

$$\|m_{P_{XY}} - m_{P_X \otimes P_Y}\|_{\mathcal{H}_X \otimes \mathcal{H}_Y}^2$$

を独立性の尺度として用いることが可能である。容易にわかるように、作用素  $\Sigma_{YX}$  の言葉で述べると、上の値は  $\Sigma_{YX}$  の Hilbert-Schmidt ノルムの 2 乗  $\|\Sigma_{YX}\|_{HS}^2$  に一致する。また、式 (11), (13) の特別な場合として、以下の表示が得られる。

$$\begin{aligned} \|\Sigma_{YX}\|_{HS}^2 &= E[k_X(X, \tilde{X}) k_Y(Y, \tilde{Y})] - 2E[E[k_X(X, \tilde{X})|X] E[k_Y(Y, \tilde{Y})|Y]] \\ &\quad + E[k_X(X, \tilde{X})] E[k_Y(Y, \tilde{Y})] \end{aligned}$$

(ただし、 $(\tilde{X}, \tilde{Y})$  は  $(X, Y)$  と独立で同一の分布に従う。)

$$\begin{aligned} \|\hat{\Sigma}_{YX}^{(n)}\|_{HS}^2 &= \frac{1}{n^2} \sum_{i,j=1}^n k_X(X_i, X_j) k_Y(Y_i, Y_j) \\ &\quad - \frac{2}{n^3} \sum_{i=1}^n \sum_{j=1}^n k_X(X_i, X_j) \sum_{\ell=1}^n k_Y(Y_i, Y_\ell) \\ &\quad + \frac{1}{n^4} \sum_{i,j=1}^n k_X(X_i, X_j) \sum_{\ell,r=1}^n k_Y(Y_\ell, Y_r) \end{aligned}$$

$\|\hat{\Sigma}_{YX}^{(n)}\|_{HS}^2$  は  $\|\Sigma_{YX}\|_{HS}^2$  に収束する。実はさらに強く、 $\|\hat{\Sigma}_{YX}^{(n)} - \Sigma_{YX}\|_{HS} = O_p(n^{-1/2})$ <sup>4</sup> が示されている (Fukumizu et al., 2007)。従って、 $\|\hat{\Sigma}_{YX}^{(n)}\|_{HS}^2$  を独立性・依存性をはかる尺度として用いることが可能である。また均一性検定と同様、独立性検定に用いることが可能であるが、その詳細については Gretton et al. (2008) を見ていただきたい。

<sup>4</sup> $O_p$  は確率的なオーダーを示す記号で、確率変数  $Z_n$  が  $Z_n = O_p(a_n)$  であるとは、任意の  $\varepsilon > 0$  に対して  $M > 0$  があって  $\limsup_{n \rightarrow \infty} \Pr(|Z_n| > M a_n) < \varepsilon$  であることと定義される。

### §3.4. カーネル法による条件付独立性的特徴づけ

相互共分散作用素を用いると、さらに確率変数の条件付独立性的特徴付けが可能となる。条件付独立性は、グラフィカルモデルをはじめとする統計的なモデリングの基礎となる重要な概念である。

#### 条件付相互共分散作用素

まず有限次元ガウス確率ベクトルの条件付共分散行列を復習する。 $X, Y, Z$  を有限次元ガウス確率ベクトルとすると、 $Z$  が与えられたもとでの  $X$  と  $Y$  の条件付共分散行列は

$$(17) \quad W_{YX|Z} = W_{YX} - W_{YZ}W_{ZZ}^{-1}W_{ZX}$$

によって定義される。ここで  $W_{YX}$  などは分散共分散行列を表し、 $W_{ZZ}$  は可逆と仮定する。よく知られているように、ガウス確率ベクトルに対しては  $\text{Cov}[Y, X|Z]$  は  $Z$  の値に依存せず、

$$(18) \quad W_{YX|Z} = \text{Cov}[Y, X|Z]$$

が成り立つ。また、 $Z$  が与えられたもとで  $X$  と  $Y$  が条件付独立であることと  $W_{YX|Z} = 0$  が同値である。

この事実を再生核ヒルベルト空間へ一種の拡張を行うことが可能であり、一般の確率変数の条件付独立性を特徴付けることができる。以下では Fukumizu et al. (2004, 2008b) に従ってその方法を説明する。

以降、 $(\mathcal{X}, \mathcal{B}_\mathcal{X}), (\mathcal{Y}, \mathcal{B}_\mathcal{Y}), (\mathcal{Z}, \mathcal{B}_\mathcal{Z})$  を可測空間、 $(X, Y, Z)$  を  $\mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$  に値をとる確率変数、 $k_\mathcal{X}, k_\mathcal{Y}, k_\mathcal{Z}$  をそれぞれ  $\mathcal{X}, \mathcal{Y}, \mathcal{Z}$  上の (A-1) を満たす可測な正定値カーネル、 $\mathcal{H}_\mathcal{X}, \mathcal{H}_\mathcal{Y}, \mathcal{H}_\mathcal{Z}$  をそれぞれ  $k_\mathcal{X}, k_\mathcal{Y}, k_\mathcal{Z}$  の定める再生核ヒルベルト空間とする。また、確率変数の周辺分布を  $P_\mathcal{X}, P_\mathcal{Y}, P_\mathcal{Z}$  で表す。

一般に  $\mathcal{H}$  をヒルベルト空間とすると、 $P_\mathcal{Z}$  に関して強 2 乗可積分な  $\mathcal{H}$ -値可測関数  $F: \mathcal{Z} \rightarrow \mathcal{H}$ 、すなわち  $E[\|F(Z)\|_\mathcal{H}^2] < \infty$  なる  $\mathcal{H}$ -値可測関数の全体は、実数値関数の場合と同様、 $\{F: \mathcal{Z} \rightarrow \mathcal{H} \mid E\|F(Z)\|^2 = 0\}$  を法としてヒルベルト空間となる。これを  $L^2(\mathcal{Z}; \mathcal{H}, P_\mathcal{Z})$  で表す。内積は  $(F, G)_{L^2(\mathcal{Z}; \mathcal{H}, P_\mathcal{Z})} = E[\langle F(Z), G(Z) \rangle_\mathcal{H}]$  で与えられる。容易に確認できるように

$$L^2(\mathcal{Z}; \mathcal{H}, P_\mathcal{Z}) \cong L^2(\mathcal{Z}, P_\mathcal{Z}) \otimes \mathcal{H}$$

という同一視が成り立つ。 $\phi_i \in L^2(\mathcal{Z}, P_\mathcal{Z})$  と  $f_i \in \mathcal{H}$  ( $i = 1, \dots, n$ ) に対して、 $[z \mapsto \sum_{i=1}^n \phi_i(z)f_i] \in L^2(\mathcal{Z}; \mathcal{H}, P_\mathcal{Z})$  と  $\sum_{i=1}^n \phi_i \otimes f_i \in L^2(\mathcal{Z}, P_\mathcal{Z}) \otimes \mathcal{H}$  が対応する。

いま、 $\mathcal{H} = \mathcal{H}_\mathcal{X}$  として  $L^2(\mathcal{Z}; \mathcal{H}_\mathcal{X}, P_\mathcal{Z})$  上の汎関数

$$L^2(\mathcal{Z}; \mathcal{H}_\mathcal{X}, P_\mathcal{Z}) \ni \Phi \quad \mapsto \quad E[\langle \Phi(Z), k_\mathcal{X}(\cdot, X) \rangle_{\mathcal{H}_\mathcal{X}}] \in \mathbb{R}$$

を定義すると、 $|E[\langle \Phi(Z), k_X(\cdot, X) \rangle_{\mathcal{H}_X}]| \leq E[\|\Phi(Z)\|_{\mathcal{H}_X} \|k(\cdot, X)\|_{\mathcal{H}_X}]$   
 $\leq \sqrt{E\|\Phi(Z)\|_{\mathcal{H}_X}^2} \sqrt{E[k_X(X, X)]}$  により、この汎関数は連続である。したがって Riesz の補題により、ある  $M_{X|Z} \in L^2(\mathcal{Z}; \mathcal{H}, P_Z)$  が存在し、任意の  $\Phi \in L_2(\mathcal{Z}; \mathcal{H}_X, P_Z)$  に対し

$$\langle M_{X|Z}, \Phi \rangle_{L^2(\mathcal{Z}; \mathcal{H}_X, P_Z)} = E[\langle \Phi(Z), k_X(\cdot, X) \rangle_{\mathcal{H}_X}]$$

が成り立つ。特に  $\Phi = \phi \otimes f$  ( $\phi \in L^2(\mathcal{Z}, P_Z), f \in \mathcal{H}_X$ ) とすると

$$(M_{X|Z}, \phi \otimes f)_{L_2(P_Z) \otimes \mathcal{H}_X} = E[\phi(Z)f(X)]$$

を得るが、これは、任意の  $\phi \in L^2(\mathcal{Z}, P_Z)$  に対して

$$(\langle M_{X|Z}, f \rangle_{\mathcal{H}_X}, \phi)_{L_2(P_Z)} = E[\phi(Z)f(X)] = (E[f(X)|Z], \phi)_{L_2(P_Z)}$$

を意味するので、次の補題が成り立つ。

**補題 3.2.** 任意の  $f \in \mathcal{H}_X$  に対して

$$(19) \quad \langle M_{X|Z}, f \rangle_{\mathcal{H}_X} = E[f(X)|Z]$$

が成り立つ。

$M_{X|Z}$  を、 $Z$  が与えられたもとでの ( $\mathcal{H}_X$  における)  $X$  の条件付平均と呼ぶ。

$M_{X|Z} \in L_2(\mathcal{Z}; \mathcal{H}_X, P_Z)$  と  $M_{Y|Z} \in L_2(\mathcal{Z}; \mathcal{H}_Y, P_Z)$  をそれぞれ  $Z$  が与えられたもとでの  $X$  と  $Y$  の条件付平均とすると、 $M_{X|Z} \otimes M_{Y|Z}$  は  $L_1(\mathcal{Z}; \mathcal{H}_X \otimes \mathcal{H}_Y, P_Z)$  の元である。この平均を

$$m_{X \perp Y|Z} \equiv E[M_{X|Z} \otimes M_{Y|Z}]$$

と表すことにすると、平均の性質 (式 6) と補題 3.2 により次の事実が成り立つ。

**補題 3.3.** 任意の  $f \in \mathcal{H}_X$  と  $g \in \mathcal{H}_Y$  に対し

$$(20) \quad \langle m_{X \perp Y|Z}, f \otimes g \rangle_{\mathcal{H}_X \otimes \mathcal{H}_Y} = E_Z[E[f(X)|Z]E[g(Y)|Z]]$$

が成り立つ。

作用素  $C_{YX|Z} : \mathcal{H}_X \rightarrow \mathcal{H}_Y$  を

$$(21) \quad \langle g, C_{YX|Z}f \rangle_{\mathcal{H}_Y} = \langle m_{XY} - m_{X \perp Y|Z}, f \otimes g \rangle_{\mathcal{H}_X \otimes \mathcal{H}_Y}$$

により定義しよう。ここで  $m_{XY}$  は  $\mathcal{H}_X \otimes \mathcal{H}_Y$  上での  $(X, Y)$  の平均である。このとき

$$\begin{aligned} \langle g, C_{YX|Z}f \rangle_{\mathcal{H}_Y} &= E[f(X)g(Y)] - E[E[f(X)|Z]E[g(Y)|Z]] \\ &= E[\text{Cov}[f(X), g(Y)|Z]] \quad (\forall f \in \mathcal{H}_X, g \in \mathcal{H}_Y) \end{aligned}$$

であるので、 $C_{YX|Z}$  を条件付相互共分散作用素と呼ぶ。

$k_Z$  が特性的であるとき、条件付相互共分散作用素は  $W_{YX|Z}$  と同様の表現を持つ。これをみるため

$$(22) \quad \Sigma_{YX|Z} \equiv \Sigma_{YX} - \Sigma_{YY}^{1/2} V_{YZ} V_{ZX} \Sigma_{XX}^{1/2}$$

と定義する。厳密性に目を瞑れば  $\Sigma_{YX|Z} = \Sigma_{YX} - \Sigma_{YZ} \Sigma_{ZZ}^{-1} \Sigma_{ZX}$  と考えてもよく、式 (18) の拡張である。このとき、次の定理が成り立つ。

**定理 3.4.**  $k_Z$  が特性的であるとき、仮定 (A-1) のもと

$$\Sigma_{YX|Z} = C_{YX|Z}$$

が成り立つ。特に任意の  $f \in \mathcal{H}_X$  と  $g \in \mathcal{H}_Y$  に対し

$$(23) \quad \langle g, \Sigma_{YX|Z} f \rangle_{\mathcal{H}_Y} = E[\text{Cov}[f(X), g(Y)|Z]]$$

である。

証明.  $\langle g, \Sigma_{YY}^{1/2} V_{YZ} V_{ZX} \Sigma_{XX}^{1/2} f \rangle_{\mathcal{H}_Y} = E[E[f(X)|Z]E[g(X)|Z]]$  を示せば十分である。 $\Sigma_{ZZ}$  は Hilbert-Schmidt 作用素なので、 $\mathcal{H}_Z$  の完全正規直交系  $\{\phi_i\}_{i=1}^\infty$  で  $\Sigma_{ZZ}\phi_i = \lambda_i\phi_i$  ( $\lambda_i \geq 0$ ) を満たすものが存在する。 $I_+ = \{i \in \mathbb{N} \mid \lambda_i > 0\}$  とおき、 $i \in I_+$  に対し

$$\tilde{\phi}_i = \frac{1}{\sqrt{\lambda_i}}(\phi_i - E[\phi(Z)])$$

と定める。 $\mathcal{R}(V_{ZY})$  と  $\mathcal{R}(V_{ZX})$  が  $\mathcal{N}(\Sigma_{ZZ})$  に直交することから、

$$\begin{aligned} \langle g, \Sigma_{YY}^{1/2} V_{YZ} V_{ZX} \Sigma_{XX}^{1/2} f \rangle_{\mathcal{H}_Y} &= \sum_{i \in I_+} \langle \phi_i, V_{ZY} \Sigma_{YY}^{1/2} g \rangle_{\mathcal{H}_Y} \langle \phi_i, V_{ZX} \Sigma_{XX}^{1/2} f \rangle_{\mathcal{H}_X} \\ &= \sum_{i \in I_+} \left\langle \frac{1}{\sqrt{\lambda_i}} \phi_i, \Sigma_{ZY} g \right\rangle_{\mathcal{H}_Y} \left\langle \frac{1}{\sqrt{\lambda_i}} \phi_i, \Sigma_{ZX} f \right\rangle_{\mathcal{H}_X} \\ &= \sum_{i \in I_+} E[\tilde{\phi}_i(Z)(g(Y) - E[g(Y)])] E[\tilde{\phi}_i(Z)(f(X) - E[f(X)])] \\ &= \sum_{i \in I_+} (\tilde{\phi}_i, E[g(Y)|Z] - E[g(Y)])_{L_2(P_Z)} (\tilde{\phi}_i, E[f(X)|Z] - E[f(X)])_{L_2(P_Z)} \end{aligned}$$

を得る。 $\{\tilde{\phi}_i\}_{i \in I_+}$  は  $L_2(P_Z)$  の正規直交系であるが、補題 3.1 から  $\{\tilde{\phi}_i\}_{i \in I_+ \cup \{1\}}$  が完全正規直交系をなす。これにより、

$$\begin{aligned} \langle g, \Sigma_{YY}^{1/2} V_{YZ} V_{ZX} \Sigma_{XX}^{1/2} f \rangle_{\mathcal{H}_Y} &= (E[g(Y)|Z] - E[g(Y)], E[f(X)|Z] - E[f(X)])_{L_2(P_Z)} \\ &\quad - (1, E[g(Y)|Z] - E[g(Y)])_{L_2(P_Z)} (1, E[f(X)|Z] - E[f(X)])_{L_2(P_Z)} \\ &= E_Z \left[ (E[g(Y)|Z] - E[g(Y)])(E[f(X)|Z] - E[f(X)]) \right] \end{aligned}$$

となり、定理が証明される。 □

定理 3.4 は、ガウス確率変数の場合の関係式 (18) の一種の拡張であるが、定理 3.4 は  $\text{Cov}[X, Y|Z]$  ではなく  $E[\text{Cov}[X, Y|Z]]$  を表現していることに注意されたい。

### 条件付独立性的特徴づけ

以下では  $Z$  が与えられたもとでの  $X$  と  $Y$  の条件付独立性を  $X \perp\!\!\!\perp Y|Z$  で表す。ガウス確率変数の場合と異なり、再生核ヒルベルト空間の場合には、特性的なカーネルを用いたとしても、 $\Sigma_{YX|Z} = O$  と  $X \perp\!\!\!\perp Y|Z$  は同値とは限らない。これは、ガウス確率変数の場合には  $\text{Cov}[Y, X|Z]$  が  $Z$  に依存しないのに対し、定理 3.4 の  $\langle g, \Sigma_{YX|Z} f \rangle$  が  $Z$  に関する期待値しか表せないことから推察できるであろう。

いま、確率変数  $(X, Y, Z)$  に対し、 $\mathcal{X} \times \mathcal{Y}$  上の確率分布  $E_Z[P_{X|Z} \otimes P_{Y|Z}]$  を、任意の  $A \in \mathcal{B}_X, B \in \mathcal{B}_Y$  に対して

$$E_Z[P_{X|Z} \otimes P_{Y|Z}](A \times B) = E_Z[E[\chi_A(X)|Z] E[\chi_B(Y)|Z]]$$

を満たすものとして定義する。また、 $(X, Y)$  の同時確率分布を  $P_{XY}$  で表すことにする。このとき、以下の定理が成り立つ。

**定理 3.5.** 正定値カーネルと確率変数は定理 3.4 と同じ仮定を満たすとし、さらに積  $k_X k_Y$  は  $\mathcal{X} \times \mathcal{Y}$  上の特性的なカーネルと仮定する。このとき

$$(24) \quad \Sigma_{YX|Z} = O \quad \Longleftrightarrow \quad P_{YX} = E_Z[P_{Y|Z} \otimes P_{X|Z}]$$

の同値関係が成立する。

証明. 定理 3.4 の表示により左向きの矢印は明らかなので、右向きを示す。簡単のため  $Q = E_Z[P_{X|Z} \otimes P_{Y|Z}]$  と書く。 $\Sigma_{YX|Z} = O$  のとき、定理 3.4 より、任意の  $f \in \mathcal{H}_X, g \in \mathcal{H}_Y$  に対して  $E_Q[f(X)g(Y)] = E_{P_{XY}}[f(X)g(Y)]$  が成り立つ。さらに、直積  $\mathcal{H}_X \otimes \mathcal{H}_Y$  の任意の元が  $\sum_{i=1}^n f_i g_i$  ( $f_i \in \mathcal{H}_X, g_i \in \mathcal{H}_Y$ ) の形の元の極限として得られることから、任意の関数  $\phi \in \mathcal{H}_X \otimes \mathcal{H}_Y$  に対して  $E_Q[\phi(X, Y)] = E_{P_{XY}}[\phi(X, Y)]$  を得る。 $k_X k_Y$  は特性的なので  $Q = P_{XY}$  が成り立つ。 □

上の定理は、 $\Sigma_{YX|Z} = O$  が  $X \perp\!\!\!\perp Y|Z$  よりも弱い条件であることを示している。しかしながら、 $X$  と  $Y$  の代わりに  $(X, Z)$  または  $(Y, Z)$  を用いると、条件付独立性的が特徴付けられる。

**定理 3.6.** 正定値カーネルと確率変数は定理 3.4 と同じ仮定を満たすとする。さらに、 $U = (X, Z)$  とし、 $\mathcal{X} \times \mathcal{Z}$  上の正定値カーネル  $k_U = k_X k_Z$  を定め



るとき、積  $k_U k_Y$  が  $(\mathcal{X} \times \mathcal{Z}) \times \mathcal{Y}$  上の特性的なカーネルであると仮定する。このとき、

$$(25) \quad \Sigma_{YU|Z} = O \quad \Longleftrightarrow \quad X \perp\!\!\!\perp Y | Z$$

の同値関係が成立する。

証明. 左方向の矢印のみ示せばよい。一般に、任意の可測集合  $A \in \mathcal{B}_X, B \in \mathcal{B}_Y, C \in \mathcal{B}_Z$  に対し、

$$\begin{aligned} & E \left[ E[\chi_{A \times C}(X, Z) | Z] E[\chi_B(Y) | Z] \right] - E \left[ \chi_{A \times C}(X, Z) \chi_B(Y) \right] \\ &= E \left[ E[\chi_A(X) | Z] \chi_C(Z) E[\chi_B(Y) | Z] \right] - E \left[ E[\chi_A(X) \chi_B(Y) | Z] \chi_C(Z) \right] \\ &= \int_C \left\{ P_{X|Z}(A|z) P_{Y|Z}(B|z) - P_{XY|Z}(A \times B|z) \right\} dP_Z(z) \end{aligned}$$

が成り立つが、定理 3.5 により、 $\Sigma_{YU|Z} = O$  という条件は、上式末行の積分が 0 であること、すなわち、 $P_{X|Z}(A|z) P_{Y|Z}(B|z) - P_{XY|Z}(A \times B|z) = 0$  が  $P_Z$  に関して確率 1 で成り立つことと同値である。これは  $X \perp\!\!\!\perp Y | Z$  を意味する。□

### 条件付相互共分散作用素の推定

条件付相互共分散作用素の推定量を構成するには

$$\Sigma_{YX|Z} = \Sigma_{YX} - \Sigma_{YZ} \Sigma_{ZZ}^{-1} \Sigma_{ZX}$$

の表現を用いて、相互共分散作用素の推定量に置き換えるのが簡単である。このとき  $\hat{\Sigma}_{ZZ}^{(n)}$  の逆作用素を考える必要があるが、これは一般に存在するとは限らないため、正則化

$$(\hat{\Sigma}_{ZZ}^{(n)} + \varepsilon_n I)^{-1}$$

( $\varepsilon_n > 0$ ) を用いる。ここで  $\varepsilon_n$  は正則化のための正定数で  $n \rightarrow \infty$  のとき  $\varepsilon_n \rightarrow 0$  となるように定める。これを用いて、条件付相互共分散作用素の推定量  $\hat{\Sigma}_{YX|Z}^{(n)}$  を

$$(26) \quad \hat{\Sigma}_{YX|Z}^{(n)} := \hat{\Sigma}_{YX}^{(n)} - \hat{\Sigma}_{YZ}^{(n)} (\hat{\Sigma}_{ZZ}^{(n)} + \varepsilon_n I)^{-1} \hat{\Sigma}_{ZX}^{(n)}$$

により定める。

$\ddot{X} = (X, Z)$  とするとき、独立性の場合と同様に、 $\hat{\Sigma}_{Y\ddot{X}|Z}^{(n)}$  の Hilbert-Schmidt ノルムを条件付独立性の尺度として用いることが可能である。さらに対称性を重視して、 $\ddot{Y} = (Y, Z)$  とおき、 $\|\hat{\Sigma}_{\ddot{X}\ddot{Y}|Z}^{(n)}\|_{HS}^2$  を用いることもできる。これを中心化グラム行列  $G_{\ddot{X}}, G_{\ddot{Y}}, G_Z$  を用いて書き下すと、

$$\begin{aligned} \|\hat{\Sigma}_{\ddot{X}\ddot{Y}|Z}^{(n)}\|_{HS}^2 &= \frac{1}{n^2} \text{Tr} \left[ G_{\ddot{X}} G_{\ddot{Y}} - 2 G_{\ddot{X}} G_Z (G_Z + n \varepsilon_n I_n)^{-1} G_{\ddot{Y}} + \right. \\ &\quad \left. G_{\ddot{X}} G_Z (G_Z + n \varepsilon_n I_n)^{-1} G_{\ddot{Y}} G_Z (G_Z + n \varepsilon_n I_n)^{-1} \right] \end{aligned}$$

となる。Fukumizu et al. (2006) の Lemma 7, 10 の証明と同様の議論を用いると、 $\varepsilon_n \rightarrow 0$  かつ  $n\varepsilon_n^3 \rightarrow \infty$  のとき  $n \rightarrow \infty$  において  $\|\hat{\Sigma}_{\tilde{X}\tilde{Y}|Z}^{(n)}\|_{HS}^2$  が  $\|\Sigma_{\tilde{X}\tilde{Y}|Z}\|_{HS}^2$  に確率収束することも示される。

Fukumizu et al. (2008a) ではこの尺度を用いた条件付独立性の検定の例が示されている。また Sun et al. (2007) では、今まで述べた Hilbert-Schmidt ノルムによる独立性及び条件付独立性の尺度を用いて、変数間の因果関係の推論を行う方法を提案している。

### §3.5. 正規化された相互共分散作用素

本節では、(条件付) 相互共分散作用素を正規化して得られる作用素を考える。正規化されていない場合と同様、特性的なカーネルを用いれば独立性や条件付独立性が特徴付けられるが、正規化された作用素が Hilbert-Schmidt であれば、その Hilbert-Schmidt ノルムはカーネルに依存しない値を持つことが示される。以下 Fukumizu et al. (2008a) にしたがってこの事実を説明する。

$(X, Y, Z)$ ,  $k_X, k_Y, k_Z$  および  $\mathcal{H}_X, \mathcal{H}_Y, \mathcal{H}_Z$  は前節と同じ意味で用いる。 $X$  と  $Y$  の相互共分散作用素

$$\Sigma_{YX} = \Sigma_{YY}^{1/2} V_{YX} \Sigma_{XX}^{1/2}$$

の分解に表れる有界作用素  $V_{YX}$  を  $X$  と  $Y$  の正規化された相互共分散作用素と呼ぶことにする。また、正規化された条件付相互共分散作用素を

$$(27) \quad V_{YX|Z} = V_{YX} - V_{YZ}V_{ZX}$$

により定義する。作用素  $V_{YX|Z}$  は、 $V_{YX|Z} = \Sigma_{YY}^{-1/2} \Sigma_{YX|Z} \Sigma_{XX}^{-1/2}$  と表現することもできる。

$\mathcal{N}(\Sigma_{XX})$  に属する関数は、 $P_X$  に関してほとんどいたるところ定数に限られるので、容易に確認できるように、定理 3.3 および 3.6 の類似が  $V_{YX}$  と  $V_{\tilde{Y}\tilde{X}|Z}$  に関して成り立ち、正規化された作用素によっても独立性および条件付独立性が特徴付けられる。

$(X_1, Y_1, Z_1), \dots, (X_n, Y_n, Z_n)$  を  $(X, Y, Z)$  と同じ分布に従う i.i.d. サンプルとすると、正規化された (条件付) 相互共分散作用素の推定量は、条件付相互共分散作用素の場合と同様にして定義することができる。すなわち、

$$\hat{V}_{YX}^{(n)} = (\hat{\Sigma}_{YY}^{(n)} + \varepsilon_n I)^{-1/2} \hat{\Sigma}_{YX}^{(n)} (\hat{\Sigma}_{XX}^{(n)} + \varepsilon_n I)^{-1/2},$$

( $\varepsilon_n > 0$  は正規化定数)

$$\hat{V}_{YX|Z}^{(n)} = \hat{V}_{YX}^{(n)} - \hat{V}_{YZ}^{(n)} \hat{V}_{ZX}^{(n)}$$

とすればよい。

正規化された (条件付) 共分散作用素は一般に Hilbert-Schmidt とは限らないが、 ${}^5\hat{V}_{YX}^{(n)}, \hat{V}_{YX|Z}^{(n)}$  は有限ランクであり Hilbert-Schmidt である。その Hilbert-Schmidt ノルムは以下のようにグラム行列を用いて計算可能である。

$$(28) \quad \|\hat{V}_{YX}^{(n)}\|_{HS}^2 = \text{Tr}[R_Y R_X],$$

$$(29) \quad \|\hat{V}_{YX|Z}^{(n)}\|_{HS}^2 = \text{Tr}[R_Y R_X - 2R_Y R_X R_Z + R_Y R_Z R_X R_Z].$$

ここで  $G_X, G_Y, G_Z$  は中心化グラム行列であり、 $R_X, R_Y, R_Z$  は

$$R_X = G_X(G_X + n\varepsilon_n I_n)^{-1}, R_Y = G_Y(G_Y + n\varepsilon_n I_n)^{-1}, R_Z = G_Z(G_Z + n\varepsilon_n I_n)^{-1}$$

により定義される。

前節で定義した (条件付) 独立性・依存性の尺度  $\|\Sigma_{YX|Z}\|_{HS}^2, \|\Sigma_{YX}\|_{HS}^2$  の値は正定値カーネルの選択に依存する。一方、特性的なカーネルに対して正規化された (条件付) 共分散作用素を用いると、それらが Hilbert-Schmidt のとき、その Hilbert-Schmidt ノルムはカーネルに依存しない値をとるという興味深い性質を持つ。

**定理 3.7.**  $\mu_X$  と  $\mu_Y$  をそれぞれ  $(\mathcal{X}, \mathcal{B}_X)$  と  $(\mathcal{Y}, \mathcal{B}_Y)$  上の測度とし、確率  $P_{XY}$  と  $E_Z[P_{X|Z} \otimes P_{Y|Z}]$  はそれぞれ  $\mu_X \times \mu_Y$  に関して確率密度関数  $p_{XY}$  と  $p_{X \perp Y|Z}$  を持つとする。 $k_Z$  と  $k_X k_Y$  を特性的なカーネルと仮定すると、 $V_{YX|Z}$  は Hilbert-Schmidt であり

$$(30) \quad \|V_{YX|Z}\|_{HS}^2 = \int \int_{\mathcal{X} \times \mathcal{Y}} \left( \frac{p_{XY}(x, y)}{p_X(x)p_Y(y)} - \frac{p_{X \perp Y|Z}(x, y)}{p_X(x)p_Y(y)} \right)^2 p_X(x)p_Y(y) d\mu_X d\mu_Y$$

が成立する。ここで  $p_X, p_Y$  はそれぞれ周辺分布  $P_X, P_Y$  の確率密度関数である。特に  $Z = \emptyset$  とすると

$$(31) \quad \|V_{YX}\|_{HS}^2 = \int \int_{\mathcal{X} \times \mathcal{Y}} \left( \frac{p_{XY}(x, y)}{p_X(x)p_Y(y)} - 1 \right)^2 p_X(x)p_Y(y) d\mu_X d\mu_Y$$

が成り立つ。

証明.  $\{\phi_i\}_{i=1}^\infty$  と  $\{\psi_j\}_{j=1}^\infty$  を、それぞれ  $\mathcal{H}_X$  と  $\mathcal{H}_Y$  の完全正規直交系で  $\Sigma_{XX}$  と  $\Sigma_{YY}$  の固有ベクトルからなるもの、すなわち  $\Sigma_{XX}\phi_i = \lambda_i\phi_i$  ( $\lambda_i \geq 0$ ),

<sup>5</sup>例えば  $X = Y$  であれば  $V_{YX}$  は単位作用素で Hilbert-Schmidt でない。

再生核ヒルベルト空間を用いた統計的推論

$\Sigma_{YY}\psi_j = \nu_j\psi_j$  ( $\nu_j \geq 0$ ) を満たすように取る。このとき

(32)

$$\begin{aligned} \sum_{i,j=1}^{\infty} \langle \psi_j, V_{YX|Z}\phi_i \rangle_{\mathcal{H}_Y}^2 &= \sum_{i,j=1}^{\infty} \langle \psi_j, V_{YX}\phi_i \rangle_{\mathcal{H}_Y}^2 \\ &\quad - 2 \sum_{i,j=1}^{\infty} \langle \psi_j, V_{YX}\phi_i \rangle_{\mathcal{H}_Y} \langle \psi_j, V_{YZ}V_{ZX}\phi_i \rangle_{\mathcal{H}_Y} + \sum_{i,j=1}^{\infty} \langle \psi_j, V_{YZ}V_{ZX}\phi_i \rangle_{\mathcal{H}_Y}^2 \end{aligned}$$

である。定理 3.4 の証明と同様に  $I_+^X = \{i \in \mathbb{N} \mid \lambda_i > 0\}$ ,  $I_+^Y = \{i \in \mathbb{N} \mid \nu_i > 0\}$  とおき、 $i \in I_+^X$ ,  $j \in I_+^Y$  に対し、 $\tilde{\phi}_i = (\phi_i - E[\phi_i(X)])/\sqrt{\lambda_i}$ ,  $\tilde{\psi}_j = (\psi_j - E[\psi_j(Y)])/\sqrt{\nu_j}$  と定める。すると、式 (32) の第 1 項は

$$\begin{aligned} \sum_{i \in I_+^X, j \in I_+^Y} \langle \tilde{\psi}_j, \Sigma_{YX}\tilde{\phi}_i \rangle_{\mathcal{H}_Y}^2 &= \sum_{i \in I_+^X, j \in I_+^Y} E_{YX}[\tilde{\psi}_j(Y)\tilde{\phi}_i(X)]^2 \\ &= \sum_{i \in I_+^X, j \in I_+^Y} \left( \int \int_{\mathcal{X} \times \mathcal{Y}} p_{XY}(x, y) \tilde{\phi}_i(x) \tilde{\psi}_j(y) d\mu_X d\mu_Y \right)^2 \end{aligned}$$

と表される。ここで  $\tilde{\phi}_0 = 1$ ,  $\tilde{\psi}_0 = 1$  と書くことにすると、 $k_X k_Y$  が特性的であることから、補題 3.1 により  $\{\tilde{\phi}_i \tilde{\psi}_j\}_{i \in I_+^X \cup \{0\}, j \in I_+^Y \cup \{0\}}$  は  $L_2(P_X \otimes P_Y)$  の完全正規直交系である。したがって、上式の末行はさらに

$$\begin{aligned} &\left\| \frac{p_{XY}(x, y)}{p_X(x)p_Y(y)} \right\|_{L_2(P_X \otimes P_Y)}^2 - \sum_{i \in I_+^X} E[\tilde{\phi}_i(X)] - \sum_{j \in I_+^Y} E[\tilde{\psi}_j(Y)] - 1 \\ &= \int \int_{\mathcal{X} \times \mathcal{Y}} \frac{p_{XY}^2(x, y)}{p_X(x)p_Y(y)} d\mu_X d\mu_Y - 1 \end{aligned}$$

と表される。

また、式 (32) の第 2 項は

$$\begin{aligned}
& \sum_{i,j=1}^{\infty} \langle \psi_j, V_{YX} \phi_i \rangle_{\mathcal{H}_Y} \langle \psi_j, V_{YZ} V_{ZX} \phi_i \rangle_{\mathcal{H}_Y} \\
&= \sum_{i \in I_+^X, j \in I_+^Y} \langle \tilde{\psi}_j, \Sigma_{YX} \tilde{\phi}_i \rangle_{\mathcal{H}_Y} \langle \tilde{\psi}_j, \Sigma_{YY}^{1/2} V_{YZ} V_{ZX} \Sigma_{XX}^{1/2} \tilde{\phi}_i \rangle_{\mathcal{H}_Y} \\
&= \sum_{i \in I_+^X, j \in I_+^Y} E[\tilde{\psi}_j(Y) \tilde{\phi}_i(X)] E[E[\tilde{\psi}_j(Y)|Z] E[\tilde{\phi}_i(X)|Z]] \\
&= \sum_{i \in I_+^X, j \in I_+^Y} \int \int_{\mathcal{X} \times \mathcal{Y}} \tilde{\psi}_j(y) \tilde{\phi}_i(x) p_{XY}(x, y) d\mu_X d\mu_Y \\
&\quad \times \int \int_{\mathcal{X} \times \mathcal{Y}} \tilde{\psi}_j(y) \tilde{\phi}_i(x) p_{X \perp Y|Z}(x, y) d\mu_X d\mu_Y \\
&= \sum_{i \in I_+^X, j \in I_+^Y} \left( \tilde{\psi}_j \tilde{\phi}_i, \frac{p_{XY}}{p_X p_Y} \right)_{L_2(P_X \otimes P_Y)} \left( \tilde{\phi}_i \tilde{\psi}_j, \frac{p_{X \perp Y|Z}}{p_X p_Y} \right)_{L_2(P_X \otimes P_Y)}
\end{aligned}$$

と表される。第 1 項の場合と同様の議論により、これは

$$-2 \left( \frac{p_{XY}}{p_X p_Y}, \frac{p_{X \perp Y|Z}}{p_X p_Y} \right)_{L_2(P_X \otimes P_Y)} + 2$$

に一致する。また、式 (32) の第 3 項は同様にして

$$\left\| \frac{p_{X \perp Y|Z}}{p_X p_Y} \right\|_{L_2(P_X \otimes P_Y)}^2 - 1$$

となる。以上をまとめて式 (30) を得る。  $\square$

定理 3.7 は、特性的なカーネルを用いると、正定値カーネルによって与えられる  $\|\hat{V}_{YX|Z}^{(n)}\|_{HS}^2$  や  $\|\hat{V}_{YX}^{(n)}\|_{HS}^2$  の値が正定値カーネルに依らない量を表すことを示している。特に、式 (31) の右辺は、Mean square contingency (Rényi, 1970) と呼ばれる古典的な依存性尺度に一致する。Fukumizu et al. (2008a) で示されているように、正則化定数が  $\varepsilon_n \rightarrow 0$  かつ  $n^3 \varepsilon_n \rightarrow \infty$  ( $n \rightarrow \infty$ ) を満たすとき、式 (28), (29) の推定量はそれぞれ  $\|V_{YX}\|_{HS}^2$ ,  $\|V_{YX|Z}\|_{HS}^2$  に確率収束する。この事実は、正定値カーネルを用いて Mean square contingency の一致推定量が構成されることを意味している。もちろん式 (31) の右辺の積分表示を用いて、確率密度関数の推定を行うことによって尺度を推定することは可能であるが、一般に確率密度関数の推定は特に多次元のデータに対して精度が低いことが知られている。正定値カーネルを用いた推定量は、確率密度関数の陽な推定を必要としない点が有利である。

式 (31) は、

$$\|V_{YX}\|_{HS}^2 = \int \left( \frac{p_{XY}(x, y)}{p_X(x)p_Y(y)} - 1 \right) p_{XY}(x, y) d\mu_X(x) d\mu_Y(y)$$

と書き直すことができる。一方、確率変数の依存性をはかる尺度として代表的なもののひとつである相互情報量は

$$MI(X, Y) = \int \int_{\mathcal{X} \times \mathcal{Y}} p_{XY}(x, y) \log \frac{p_{XY}(x, y)}{p_X(x)p_Y(y)} d\mu_X(x) d\mu_Y(y)$$

により定義される。相互情報量も Mean square contingency もともに非負の値をとり、0 となるのは  $X$  と  $Y$  が独立な場合に限る。 $\log z \leq z - 1$  により、定理 3.7 の仮定のもと  $MI(X, Y) \leq \|V_{YX}\|_{HS}^2$  が成り立つことがわかる。

#### §4. おわりに

本稿では、機械学習や統計的学習理論の分野で近年研究が盛んとなった「カーネル法」の概要をごく簡単に述べ、著者がたずさわっている、カーネル法による確率分布に関する統計的推論についての最近の研究について述べた。

さまざまなデータ解析の手法がカーネル法の方法論に基づいて提案されたが、それらの統計的な考察は必ずしも進んでいないのが現状である。手法の理論的解析やそれにもとづく改良は今後の重要な課題であり、ヒルベルト空間上の作用素の推定理論やヒルベルト空間に値を持つ確率変数の分布論など、数学的考察に基づく研究が発展することが期待される。

#### References

- Nachman Aronszajn. Theory of reproducing kernels. *Transactions of the American Mathematical Society*, 68(3):337–404, 1950.
- Francis R. Bach and Michael I. Jordan. Kernel independent component analysis. *Journal of Machine Learning Research*, 3:1–48, 2002.
- Charles R. Baker. Joint measures and cross-covariance operators. *Transactions of the American Mathematical Society*, 186:273–289, 1973.
- Bernhard E. Boser, Isabelle M. Guyon, and Vladimir N. Vapnik. A training algorithm for optimal margin classifiers. In D. Haussler, editor, *Fifth Annual ACM Workshop on Computational Learning Theory*, pages 144–152, Pittsburgh, PA, 1992. ACM Press.
- Kenji Fukumizu, Francis R. Bach, and Arthur Gretton. Statistical consistency of kernel canonical correlation analysis. *Journal of Machine Learning Research*, 8:361–383, 2007.

- Kenji Fukumizu, Francis R. Bach, and Michael I. Jordan. Dimensionality reduction for supervised learning with reproducing kernel Hilbert spaces. *Journal of Machine Learning Research*, 5:73–99, 2004.
- Kenji Fukumizu, Arthur Gretton, Xiaohai Sun, and Bernhard Schölkopf. Kernel measures of conditional dependence. In *Advances in Neural Information Processing Systems*, volume 21, pages 489–496. MIT Press, 2008a.
- Kenji Fukumizu, Francis R. Bach, and Michael I. Jordan. Kernel dimension reduction in regression. Technical Report 715, Department of Statistics, University of California, Berkeley, 2006.
- Kenji Fukumizu, Francis R. Bach, and Michael I. Jordan. Kernel dimension reduction in regression. *The Annals of Statistics*, to appear, 2008b.
- Arthur Gretton, Karsten M. Borgwardt, Malte Rasch, Bernhard Schölkopf, and Alex Smola. A kernel method for the two-sample-problem. In B. Schölkopf, J. Platt, and T. Hoffmann, editors, *Advances in Neural Information Processing Systems 19*. MIT Press, Cambridge, MA, 2007.
- Arthur Gretton, Kenji Fukumizu, Choon Hui Teo, Le Song, Bernhard Schölkopf, and Alex Smola. A kernel statistical test of independence. In *Advances in Neural Information Processing Systems 20*, pages 585–592. MIT Press, 2008.
- Alfréd Rényi. *Probability Theory*. North-Holland, 1970.
- Bernhard Schölkopf, Alexander Smola, and Klaus-Robert Müller. Nonlinear component analysis as a kernel eigenvalue problem. *Neural Computation*, 10: 1299–1319, 1998.
- Bernhard Schölkopf and Alexander J. Smola. *Learning with Kernels*. MIT Press, 2002.
- Bharath Sriperumbudur, Arthur Gretton, Kenji Fukumizu, Gert Lanckriet, and Bernhard Schölkopf. Injective hilbert space embeddings of probability measures. In *Proceedings of the 21st Annual Conference on Learning Theory (COLT 2008)*, 2008.
- Xiaohai Sun, Dominik Janzing, Bernhard Schölkopf, and Kenji Fukumizu. A kernel-based causal learning algorithm. In *Proceedings of the 24th International Conference on Machine Learning*, pages 855–862, 2007.

Vladimir N. Vapnik. *Statistical Learning Theory*. Wiley-Interscience, 1998.

福水 健次. 正定値カーネルによる統計的推論の方法. **21 世紀の統計科学 II 自然・生物・健康の統計科学**, pages 257–290. 東京大学出版会, 2008.